IMPROVED NEGATIVE DIFFERENTIAL RESISTANCE FIELD EFFECT TRANSISTOR (NDR-FET) & CIRCUITS USING THE SAME

CROSS REFERENCE TO RELATED APPLICATIONS

The present application is a continuation-in-part of each of the following applications, all of which were filed June 22, 2000 and are hereby incorporated by reference as if fully set forth herein:

Serial no. 09/603,101 entitled "A CMOS-PROCESS COMPATIBLE, TUNABLE NDR (NEGATIVE DIFFERENTIAL RESISTANCE) DEVICE AND METHOD OF OPERATING SAME"; and

Serial No. 09/603,102 entitled " CHARGE TRAPPING DEVICE AND METHOD FOR IMPLEMENTING A TRANSISTOR HAVING A NEGATIVE DIFFERENTIAL RESISTANCE MODE"; and

Serial No. 09/602,658 entitled "CMOS COMPATIBLE PROCESS FOR MAKING A TUNABLE NEGATIVE DIFFERENTIAL RESISTANCE (NDR) DEVICE."

The present application is also related to the following applications, all of which are filed simultaneously herewith, and which are hereby incorporated by reference as if fully set forth herein:

An application entitled "INSULATED-GATE FIELD EFFECT TRANSISTOR INTEGRATED WITH NEGATIVE DIFFERENTIAL RESISTANCE (NDR) FET;" Attorney Docket No. PROG 2001 – 1; and

An application entitled "MEMORY CELL UTILIZING NEGATIVE DIFFERENTIAL RESISTANCE FIELD-EFFECT TRANSISTORS"; Attorney Docket No. PROG 2001 – 2; and

An application entitled " DUAL MODE FET & LOGIC CIRCUIT HAVING NEGATIVE DIFFERENTIAL RESISTANCE MODE"; Attorney Docket No. PROG 2001 – 3.

An application entitled "CHARGE PUMP FOR NEGATIVE DIFFERENTIAL RESISTANCE TRANSISTOR"; Attorney Docket No. PROG 2001 – 4.

FIELD OF THE INVENTION

This invention relates to semiconductor devices and more particularly to an improved negative differential resistance (NDR) FET and circuits that utilize the same. The present invention is applicable to a wide range of semiconductor integrated circuits, particularly for applications where it is desirable to integrate NDR FET devices with conventional FETs and other similar logic/memory circuits, including in SOI and memory applications.

BACKGROUND OF THE INVENTION

A new type of CMOS compatible, NDR capable FET is described in the aforementioned applications to *King et al.* referenced above. The advantages of such device are well set out in such materials, and are not repeated here.

In preferred embodiments, this device typically uses a dielectric layer for creating a charge trapping region that rapidly traps/detraps charge carriers. A number of different techniques are explained for forming said traps to achieve a desired NDR effect. It is apparent, nonetheless, that additional processing technique (and/or more optimized versions of the processes described in King et al) would be beneficial for expanding the availability fo such devices.

A current trend also is to use so called silicon-on-insulator substrates to manufacture integrated circuits. It is expected that this technology will experience rapid growth in the years to come, but to date, only two terminal NDR diodes have been implemented in such environments. Thus, there is clearly a need for an NDR device that is as easy to integrate as a conventional FET in such technology.

Another growing trend is the use of NDR devices as load elements in SRAM memory cells and other circuit applications. To date, such NDR devices have been limited to two terminal, diode type structures which have operational limitations as well as integration complexities with CMOS processing. Furthermore, it is not possible, for example, to implement a low power memory cell using a single channel technology; current approaches are limited to conventional CMOS, where both p and n type transistors are required. Accordingly, there is an apparent compelling need for a low cost, easily integrable NDR solution for such applications as well.

SUMMARY OF THE INVENTION

Accordingly, an object of the present invention is to overcome the aforementioned deficiencies in the prior art;

Another object of the present invention is to provide an NDR FET that includes additional variations and improvements over the NDR FET described in *King et. al.*;

Still a further object of the present invention is to provide an additional type of trapping layer, and/or new types of charge traps that can be used advantageously in an NDR FET;

Another object of the present invention is to provide an NDR FET that is embodied within an SOI substrate;

Still a further object of the present invention is to provide a new type of general low power, single channel technology for effectuating logic and memory circuits;

Yet another object of the present invention is to provide an improved type of NDR device that is more flexible and more easily integrated (than prior NDR diode devices) into conventional semiconductor circuits, including SRAM memory cells.

These and other objects are provided by a first aspect of the present invention, which includes a semiconductor structure comprising a semiconductor substrate, and a dielectric layer (gate insulation layer) located on the semiconductor substrate, such that an interface region is formed between the semiconductor substrate and the dielectric layer. A plurality of carrier trapping sites within the interface region are configured for trapping carriers that are electrically biased by an electrical control field to move from a channel into the interface region. Thus, a current in the channel varies from a first current value associated with a conducting condition, to a second current value associated with a non-conducting condition, where the second current value is substantially less than the first current value.

In a preferred embodiment, a trap energy level for the trapping sites in the interface region is higher than a conduction band edge of the channel. Furthermore, the trap energy level is set so that said trapping sites trap primarily hot carriers (and not normal carriers) flowing in the channel to avoid interfering with the operation of the FET. To achieve this result, a trap energy level is set to approximately .5 eV higher than the conduction band edge. The semiconductor structure is incorporated as part of an insulated gate field effect

transistor which otherwise behaves like a conventional FET in a first region of operation, but yet has NDR capability in a second region of operation.

In the preferred embodiment, the hot carriers tunnel from the channel to the trapping sites, but they are not energized to tunnel from the channel to a conduction band of the interface region. Nor is the interface region required to have a matching conduction band to facilitate a tunneling process, as required in conventional NDR devices.

Further in a preferred embodiment, an NDR FET shares one or more common structures with a conventional insulated gate field effect transistor (IGFET), so that a common set of processing operations can be used to manufacture both types of elements for an integrated circuit.

In other variations, the trapping sites can include water based traps created by a steam ambient. The NDR FET uses an n-type channel implanted with a p-type dopant so that a relatively large electric bias field can be set up to facilitate moving said carriers from said channel to said trapping sites.

In another aspect of the invention, a memory cell includes at least one first dopant type channel insulated gate field effect transistor (IGFET). The first-channel type IGFET has an IGFET gate terminal, an IGFET source terminal connected to a first potential, and an IGFET drain terminal coupled to a storage node. In lieu of a conventional two terminal diode, the present invention incorporates a negative differential resistance field-effect transistor (NDR-FET) element that also has a first dopant-type channel, and acts as a pull up or pull down device when connected in series with the IGFET. The NDR FET element includes a firstNDR FET drain terminal connected to a second potential, a second NDR source terminal connected to the storage node, and a third NDR gate terminal connected to a bias voltage. In this fashion the memory cell is formed entirely of active devices having a common channel dopant type.

In a preferred embodiment, the NDR FET element and the IGFET share at least a common substrate and a common gate insulation layer. In addition, a common gate terminal for both can be fabricated from a single conductive layer. The two devices can further share one or more source/drain regions.

In this fashion, an NDR memory cell can be constructed that is integrated into a conventional fabrication process much easier than conventional NDR diodes. Furthermore, the cell can be made so that both devices use single channel type of dopant (i.e., both are n-

channel or p-channel), and yet still achieve low power operation as with CMOS implementations.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is the schematic cross-sectional view of an embodiment of the NDR metal-insulator-semiconductor field-effect transistor (MISFET) disclosed in this invention, including a body contact terminal 125 for receiving an NDR mode enable/disable bias signal;

Figure 2 is a graphical chart illustrating the current versus voltage (*I-V*) characteristics of the NDR-MISFET, including an NDR operating region.

Figure 3 is the schematic cross-sectional view of another embodiment of the NDR-MISFET disclosed in this invention including a body contact terminal 125 for receiving an NDR mode enable/disable bias signal;

Figure 4 is an illustrative process sequence for integrating the NDR-MISFET into a conventional CMOS logic process flow;

Figure 5 is a circuit diagram of two (2) transistor SRAM cell formed by a combination of an NDR FET and a conventional FET.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

The following detailed description is meant to be illustrative only of particular embodiments of the invention. Other embodiments of the invention and variations of those disclosed will be obvious to those skilled in the art in view of the following description.

As discussed below, a preferred device embodiment is described first. Next, the mechanism responsible for the negative differential resistance (NDR) mode is described, followed by additional preferred embodiments for enhancing the performance of an NDR device. Finally, an exemplary method of fabrication will be described.

In accordance with a preferred embodiment of the invention, an n-channel MISFET NDR device structure (Figure 1) 100 is provided which is made with minimum modification to a standard CMOS process. In fact, from a first glance, device 100 appears to be an ordinary n-channel MOS (NMOS) transistor, in which a gate electrode 110 of the device is formed on top of a semiconductor substrate 120 and electrically insulated from the substrate

by a dielectric layer 130. Right away it can be seen that NDR device 100 in this invention is distinctly different from NDR devices in the prior art.

Prior-art NDR devices are typically two-terminal diode devices, made with very complicated and expensive process sequences which are incompatible with a conventional CMOS process. Although NDR device 100 in this invention is similar in appearance to an NMOS transistor, it incorporates slight but critical modifications, as taught in this invention, in order for the device to manifest the desired NDR output characteristic mode.

A first modification is that a p-type dopant concentration in a surface region of the semiconductor substrate underneath the gate electrode (the channel) is relatively high compared to a contemporary conventionally processed n-channel device. In a preferred embodiment of device 100, the p-type dopant concentration is greater than $1 \times 10^{18}\ cm^{-3}$ in the channel. Of course, it will be understood that for any particular design rule, device characteristic and process environment the p-type dopant concentration may be varied accordingly, and that some routine design, simulation and/or testing may be necessary to optimize the performance of the device in any particular application. Accordingly, the present invention is not limited to any particular concentration, but, instead, is guided more by considerations of whether a sufficient dopant concentration has been introduced to help contribute to the NDR effect. More heavily doped n-type regions in the semiconductor surface region, adjacent to the channel and located at each end of the gate electrode, form the source and drain contact regions 140 and 150 respectively. The electric potential of the channel can be further adjusted via a body contact terminal 125.

A second modification of present device 100 over a conventional transistor is the fact that charge traps or storage nodes 135 exist in insulating layer 130 between semiconductor substrate 120 and gate electrode 110. These charge traps are located relatively close to (within 1.5 nm of) semiconductor-insulator interface 138, so that charges from semiconductor 120 can be trapped and de-trapped very quickly. Again it will be understood that this distance figure is based on the details of the present embodiment, and that for any particular environment this parameter may vary significantly, so the present invention is not limited by the particular details of the same. The key point, of course, is the existence of these charge traps, or some other physical feature that acts to store electrons. It will be understood of course that the drawing of FIG. 1 is merely an illustration to better describe the features of the present invention, and thus the arrangement and location of the

trapping sites 135 is not drawn to scale.A third modification is that insulating layer 130 between semiconductor substrate 120 and gate electrode 110 is relatively thick (greater than 6 nm) to prevent significant loss of trapped charge to the gate electrode via tunneling. Those skilled in the art will again appreciate that this thickness is again a function of the particular material, processing environment, etc., and that the present invention is by no means limited to such figure.

With source and body terminals 145 and 125 of device 100 held at ground potential and gate terminal 115 biased sufficiently high to turn on the device, the output characteristic (drain current as a function of drain voltage) of device 100 will exhibit negative differential resistance over a range of drain voltages. This aspect of the invention is illustrated in Figure 2, where device drain current versus drain voltage is plotted for two different gate voltages to show how the NDR mode can be affected by a suitable selection of the gate voltage. It can be seen that for a fixed gate voltage V_{GS}, drain current I_{DS} firstly increases in a first region 210 with drain voltage V_{DS}, similarly to the behavior that is seen in drain current in a conventional NMOS transistor. Surprisingly, however, in region 220, beyond a certain drain voltage level, drain current decreases with further increases in voltage, *i.e.* the device exhibits an NDR mode with NDR characteristics. The drain voltage at which the drain current begins to decrease (i.e., point 225 where $V_{DS} = V_{NDR}$) is adjustable through suitable selections of channel length, threshold voltage, etc. It should be noted that, due to the relatively high channel dopant concentration and the relatively thick gate dielectric, the threshold voltage of the NDR FET will be significantly higher than that of a conventional MOSFET, so that a larger than typical gate voltage is correspondingly used for the NDR FET. As a result, $V_{GS} > V_{NDR}$ so that the vertical electric field is in the direction such that electrons are attracted towards the gate electrode, enhancing the NDR effect.

This behavior by device 100 of the present invention is rather surprising, and is apparently the result of physical mechanisms that have hitherto not been exploited in this area of semiconductor devices and processing. In the prior art, band-to-band quantum-mechanical tunneling of charged particles (electrons and/or holes) from one side of a diode to the other side is known to be the primary mechanism for NDR in tunneling diodes. In contrast, for device 100 of the present invention, the physical mechanism appears to be rapid trapping of electrons in the gate insulator underneath the gate electrode, near to (within 1.5 nm of) the semiconductor-insulator interface. Referring to the device structure

in Figure 1, when device 100 is biased with a sufficiently high gate voltage such that the channel of the device is in the strong-inversion condition (*i.e.* when the gate-to-source voltage is greater than the threshold voltage), a current flows between the source and drain terminals 145 and 155 respectively of the device if a small voltage is applied between such terminals. Since the channel is configured to contain a relatively high p-type dopant concentration, a vertical (in the direction perpendicular to the semiconductor surface) electric field in the channel is large (greater than 10^6 V/cm). As the drain-to-source voltage increases, the lateral (in the direction parallel to the semiconductor surface) electric field increases, so that a composite (horizontal + vertical) electric field exerting force on inversion-layer electrons in the channel increases. Once this composite electric field reaches a certain critical value (which of course will be a function of the doping and geometry of the device) electrons flowing from source 140 to drain 150 will gain sufficient energy between collisions to surmount a semiconductor-insulator interface potential barrier. Since the vertical electric field component attracts the electrons toward gate electrode 110, electrons enter insulator 130 and subsequently are captured by the traps or storage nodes 135 in the insulator. The presence and accumulation of negative charge in insulator 130 dynamically increases a threshold voltage of device 100. In other words, the electrons accumulated in the traps/storage nodes 135 operate to set up a counter field that inhibits the movement of additional electrons into the channel from the source, and reducing an available channel current by reducing a density of electrons in the channel region. Thus, the net effect created by the traps/storage nodes 135 of the present invention is a drastic reduction in the inversion-layer charge density and commensurate reduction in the current flowing between the source and the drain. It can be seen plainly that the amount of net current in the channel that can be affected by the traps is a function of their number, concentration, location, and the bias conditions imposed on device 100, all of which are easily controllable and optimizable for any particular environment, so that the onset conditions, strength and operating region for a negative differential resistance mode can be tailored and customized as needed.

It is noted that the present disclosure teaches that only a single species of energetic carriers (hot electrons) are generated in a channel region and trapped in insulator 130, and both of these phenomena preferably occur in a substantially uniform manner throughout the channel length. This operation, too, is distinctly different from the case for a conventional

NMOS transistor, in which hot electrons are generally generated in the depletion region of the drain p-n junction, leading to impact ionization and an avalanche effect resulting in significant numbers of hot holes as well as hot electrons. Typically, this effect is maximized at a gate-to-source voltage which is lower than the drain-to-source voltage (for example, at a gate voltage equal to one half the drain voltage); hence in a conventional device the vertical electric field in the channel near the drain junction attracts hot holes, rather than hot electrons, toward the gate electrode. Clearly, then, this explains why the creation of hot electrons in a conventional NMOS transistor (even if it occurs incidentally) cannot produce the negative differential resistance characteristic as described in this invention. Furthermore it is well known that the injection of hot holes into the gate insulator causes damage, adversely affecting the performance and reliability of the NMOS transistor. In the NDR-MISFET 100 of the present invention, although holes are generated by impact ionization in the channel, they are not injected (or their injection is substantially eliminated to the point where it is negligible from an operational perspective) into gate insulator 130 because the vertical electric field repels holes from gate electrode 110.

As a point of further clarification, the mechanism responsible for the NDR characteristic of the present invention also does not require that NDR MISFET 100 be operating in a conventional "pinch-off" condition, i.e., in which a gate-to-drain voltage is lower than a threshold voltage so that the inversion-layer charge density in the channel adjacent to the drain is zero. In the pinch-off condition, the lateral electric field is non-uniformly distributed in the channel between the source and drain: the electric field increases gradually and linearly with distance away from the source, and then increases exponentially in the depletion region of the drain junction, so that the generation of hot electrons occurs predominantly in the depletion region of the drain junction, resulting in drain avalanche. In contrast, in the present invention, NDR-MISFET 100 is preferably operated in a "triode" region, so that the electric field increases uniformly from the source end of the channel to the drain end. The drain current saturates due to velocity saturation, not pinch-off, so the current does not increase linearly with V_{DS} (as seen generally in FIG.2).

In a preferred embodiment of NDR-MISFET 100, sufficient bias is applied so that the electrons in the channel become so energetic that channel hot electrons are created due to the high composite electric field in the channel. These channel hot electrons have sufficient energy imparted from the horizontal component of this field to surmount the

potential barrier at the semiconductor-insulator interface and enter gate insulator 130 because the vertical electric field component attracts them toward gate electrode 110. The electrons are captured by the traps or storage nodes 135 in insulator 130; consequently the threshold voltage of the transistor increases dynamically. More charge is trapped as the drain-to-source voltage increases (for a constant gate voltage), because the generation of hot carriers (and thus the percentage of the current that is based on a hot carrier component) correspondingly increases, and it is these hot carriers that are trapped. As greater numbers of hot carriers are trapped, they increase the threshold voltage and thereby reduce the mobile charge density in the channel by a disproportionate amount (compared to the hot-carrier current charge amount), thus decreasing the drain current dramatically. This results in the negative differential resistance in the output (drain current versus drain voltage) characteristic. It can be seen also that more charge can be trapped by increasing the vertical component of the field as well, since this increases the likelihood that a charged carrier will be forced into a trap 135 in dielectric layer 130 (the trapping rate), and also increases a temporary storage/trapping time associated with the charge. It is not necessary, nonetheless, to trap a significant number of carriers, because even a small quantity stored in the trapping sites can be sufficient to deplete the channel of mobile carriers. It is also preferable to not increase the vertical field to the point where some deleterious side effects (dielectric breakdown or lack of fast reversibility of the NDR effect for example) are seen. In other words, it is generally desirable to have the charges rapidly trapped and de-trapped at a particular rate that ensures that the device can be put into and out of an NDR mode or operating region quickly, instead of being confined to working within a particular region. Other techniques for increasing the amount of trapped charges, and the trapping/detrapping rates will be apparent to those skilled in the art. For instance, it may not be necessary in fact in some applications, to make the electrons "hot" because they will still be swept by the vertical field into the trapping sites.

Thus, the present invention uses an approach that is in contrast to that of prior art which has charge traps, such as U.S. Patent No. 5,633,178. In the prior art, the emphasis has been on retaining the charge as long as possible, and this reference for example specifically discloses using a refresh operation to keep the logic state. Accordingly, there is no effort made in the prior art to implement or sustain a dynamic process where charges are continually trapped and de-trapped. In fact conventional disclosures discourage such

condition because it has been perceived to date as an undesirable situation, and so this explains, too, why such references do not describe configuring a FET channel to have a structure and doping characteristics that would facilitate this type of trapping/detrapping mechanism.

The drain current and therefore the negative differential resistance in this invention can be adjusted by varying the gate voltage as seen in FIG. 2. As seen also in FIG. 2, the invention can be seen as exploiting the fact that, as the threshold voltage V_t dynamically increases (because of the accumulation of trapped charges) with increasing drain-to-source voltage V_{DS}, a drain current I_{DS} (which is proportional to $V_g - V_t$) will first increase, and then begin to decrease as V_t begins to exceed V_g and thus dominate the behavior of the device. Thus, a current value depicted in curve 228 will generally follow the set of continuous curves 229 shown in FIG. 2 for a given V_g and varying V_t. The so-called "peak-to-valley ratio," a key figure of merit in NDR devices, can also be tuned in the present invention through suitable combinations of doping concentrations, device geometries and applied voltages.

The present invention bears some resemblance to a leaky (or volatile) floating gate storage device. However, the trapping and de-trapping of electrons in gate insulator 130 of NDR-MISFET 100 are very rapid processes, as compared to the programming and erase processes of a conventional floating-gate non-volatile memory device, so that the threshold voltage of NDR-MISFET 100 can respond dynamically to changes in a gate-to-source voltage and/or a drain-to-source voltage. In fact, while conventional memory devices require extensive pre-programming and erase cycle times to change threshold states, the threshold voltage of the present device responds to the applied source to drain bias voltage with minimal delay. Thus, it can change and reverse a threshold (and thus achieve an NDR mode) in substantially the same time as it takes for device 100 to turn the channel on or off in response to such bias conditions. For any given bias condition (fixed gate-to-source and drain-to-source voltages), a steady-state condition exists in which electrons are continually being rapidly trapped, stored, and de-trapped, maintaining a fixed amount of net charge trapped in gate insulator 130. The fixed amount of net charge trapped in the gate insulator is dependent on the particular voltage bias conditions applied to device 100. As the gate-to-source voltage and/or the drain-to-source voltage changes, the balance of the trapping and de-trapping processes changes, thereby changing the fixed amount of net charge trapped in

the gate insulator and dynamically changing the threshold voltage. This means the net NDR effect can be controlled through two different bias parameters, a significant advantage again over conventional two terminal NDR devices. Furthermore, the negative differential resistance characteristic is seen not only as the drain-to-source voltage is increased from zero Volts to a high value (such that hot electrons are trapped in gate insulator 130), but also in the reverse direction as the drain-to-source voltage is decreased from a high value to zero Volts. It is expected, in fact that the threshold voltage variability/reversibility can be tailored to be relatively symmetric, so that it can thus be adjusted from a relatively low voltage value to a relatively high voltage value in approximately the same time required to adjust the threshold voltage from a relatively high voltage value to a relatively low voltage value.

As intimated above, the inventors believe that at higher drain to source voltages another feature of the present invention will be apparent, and that is the relatively high percentage of hot carriers in the channel current. Namely, since hot carriers are generated at a faster rate as the drain to source voltage increases the inventors believe that the net result is that eventually the hot carrier current component of the channel current will become dominant, and thus eventually constitute the only current component in the channel, even if it is extremely small overall. The relative percentage of hot carriers in the channel current, therefore, can be controlled, and this feature of the invention may be beneficial in other application environments.

Another aspect of the invention that is potentially useful is the fact that the trapping sites of the present invention can be thought of as introducing a form of current/charge delay on a single channel basis. The trapping time, temporary storage time, and detrapping time making up such delay can be controlled as a function of the applied horizontal and vertical electrical fields, and this aspect might be exploited in other environments.

As explained herein, the p-type dopant concentration in the surface region of the semiconductor underneath the gate electrode should be relatively high. This is to ensure that the vertical electric field is high (greater than 10^6 V/cm) when the transistor is turned on, to promote the creation of hot electrons in the channel. A conventional NMOS transistor with channel length less than 250 nm may (in some applications) have such a high channel dopant concentration, but it will not achieve the results of the present invention because this structure alone is insufficient to bring about an NDR effect. In a preferred embodiment, the

doping concentration is made slightly graded, so that the concentration of dopant is slightly lower at the semiconductor surface, and then peaks at some relatively small distance (below 30nm) below the surface. This is done in order to achieve a built-in electric field, which in turn serves to confine electrons near the surface of the semiconductor, and thus further enhances the injection of electrons into the trapping sites in the dielectric. Again, other doping concentrations and techniques can also be employed to induce this same phenomenon.

Furthermore, to minimize the possibility of drain avalanche, a preferred embodiment herein teaches that the drain dopant-concentration profile at the junction with the channel is made to be relatively lightly doped. This not only minimizes the impact ionization current between the drain and the channel, but also has the side benefit of minimizing the capacitance between them. By minimizing the drain junction capacitance to the channel, the overall device switching performance is enhanced and the device thus operates faster. Those skilled in the art will appreciate that there are other ways to enhance the generation of hot electrons in the channel in addition to those described herein, and the present invention is not limited to any particular implementation of the same.

A preferred embodiment also confines the relatively high dopant concentration in the channel to the surface region only, so that the dopant concentration in the channel region is initially low (to confine electrons to the surface region), then increases, and then is made lower away from the surface to achieve the effect of low drain-junction capacitance. As alluded to earlier, the present invention is not limited to any particular doping concentration and profile of the dopant in the channel, because the range of such parameters necessary to bring about the NDR effect will vary from device to device of course, depending on the size, geometry, intended function, etc., of the device, but these details can be gleaned with routine and conventional simulation and testings for any particular application, in the same manner as is done for any other conventional semiconductor device. As explained previously, the high surface dopant concentration in the channel should also be offset from the highest dopant concentration in drain region 150 through the use of lightly doped drain (LDD) structures.

One additional and very desirable feature of the present invention is that the drain voltage at the onset of negative differential resistance can be scaled with the scaling of the CMOS technology. In other words, as the transistor channel length is reduced, the drain

voltage required to reach the critical composite electric field in the channel (corresponding to the onset of negative differential resistance) is commensurately reduced. This aspect of the invention ensures that the structures and methods taught herein are guaranteed to have substantial and meaningful future utility in advanced generations of devices and products that are made using smaller geometries, lower bias conditions, etc. than those currently available.

As is evident, a key feature of NDR-MISFET 100 is that charge traps or storage nodes 135 exist in gate insulator 130, very near to (within 1.5 nm of) the semiconductor-insulator interface, so that electrons can be trapped and de-trapped very quickly. The creation and distribution/location of such traps 135 can be accomplished in any number of ways that are compatible with conventional semiconductor processing techniques. For example, traps 135 can consist of defect sites within gate dielectric 130 as shown in Figure 1, or interfacial traps 135 between two or more layers of a multi-layered gate-insulator stack, or one or more electrically isolated ("floating") conductor or semiconductor electrodes 137 embedded within a gate insulator 130 (made up of two layers 130' and 130" sandwiching the embedded electrode 137) as shown in Figure 3. The only important consideration is that the carrier trapping sites are configured for trapping carriers that are electrically biased by an electrical control field (i.e., the combined effect of bias conditions resulting from the channel doping, the gate to source voltage, the source to drain voltage) to move from the channel into insulator/dielectric layer 130. This can be done in any number of different concentrations and arrangements within layer 130 so that the channel current can be varied all the way from essentially zero (no conduction) to full conduction in accordance with the strength of the electrical control field.

In a preferred embodiment of the present invention, Boron atoms incorporated into gate insulator 130 during a thermal oxidation of heavily boron-doped silicon serve to provide defect sites which readily trap charge. Alternative embodiments may employ alternative dopant species such as Indium to form charge traps 135, and the present invention is not limited to any particular dopant species in this regard.

As mentioned, other possible embodiments may employ a multi-layered gate insulator, for example a very thin interfacial layer of silicon dioxide and a thicker layer of a second dielectric material such as silicon nitride, with charge-trapping sites at the dielectric-dielectric interface. Further possible embodiments may incorporate islands of metal, silicon

or germanium nanocrystals embedded within gate insulator, or perhaps even a single continuous floating gate electrode (Figure 3) 137, to trap charge. In fact, the present approach can be taken to an extreme to effectuate a new type of non-volatile floating gate electrode for a flash memory cell. It can be seen that complete non-volatility can be achieved by simply locating the trapping sites sufficiently far away from the interface so that the charge does not leak off after it is put there (using conventional programming techniques). This type of discontinuous floating gate electrode, formed as a multitude of trapping sites distributed in the gate dielectric, may have significant operating advantages over conventional continuous electrode. In particular, in the distributed charge storage sites aspect of the present invention, the trapped charge has less mobility than an electron in a sheet type electrode, and thus the charge storage sites are less likely to leak the stored charge (individually and in the aggregate of course) to the source/drain regions. This in turn means that the charge storage sites can be located closer to the channel, and thus the gate insulating layer can be thinner, the programming voltage and/or current smaller, etc., Other methods and techniques for creating and distributing traps 135 in a fashion suitable for achieving an NDR effect, and any non-volatile effects as shown herein will be apparent to those skilled in the art from the present teachings, and can be further gleaned from the descriptions given in the aforementioned prior art references for creating different types and arrangements of charge traps.

To enhance the electron trapping stemming from the generation of hot electrons in the channel (since it is the primary mechanism responsible for the negative differential resistance characteristic) the present disclosure also teaches a preferred embodiment of an insulator 130 for retaining the trapped charge under high gate-voltage bias. To avoid the loss of trapped electrons to gate electrode 110 via tunneling through gate insulator 130, the latter should have sufficient thickness to prevent or at least substantially reduce such tunneling effects. In a preferred embodiment insulator 130 is silicon dioxide formed by either one of, or a combination of conventional thermal oxidation and deposition techniques. As referred to earlier, to avoid significant loss of trapped charge due to quantum-mechanical tunneling, gate insulator 130 is formed to have a thickness of at least 6 nm. Other implementations of insulator material for layer 130 include Silicon Nitride (Si_3N_4), or Silicon Oxynitride (SiO_xN_y), or a high-permittivity dielectric (relative permittivity greater than 8). The use of a high-permittivity gate dielectric is advantageous for achieving high areal gate

capacitance, which facilitates adequate gate control of the channel potential. Again, the present invention is not restricted to any particular selection of thickness and material for insulator layer 130, and other variations/techniques for achieving a reduction in quantum-mechanical tunnelling known in the art can be used to the extent they are compatible with the present objectives.

For a preferred embodiment of this invention, polycrystalline silicon (poly-Si) is used as the material for gate-electrode 110. Other possible embodiments may utilize alternative gate materials such as polycrystalline silicon-germanium or metals, or any number of other conventional materials.

An exemplary process for fabricating the NDR-MISFET in a conventional CMOS fabrication facility is depicted in FIG. 4 A standard p-type silicon starting substrate 120 is first processed through standard isolation-structure-formation process steps; the surface of substrate 120 is then moderately doped (to $\sim 5\times10^{18}$ cm^{-3}) by a shallow Boron implant. Subsequent to this a deposition of silicon dioxide (~6 nm) is done (or thermal oxidation) in a manner so that the Boron becomes incorporated into a gate insulator 130 near the surface of silicon substrate 120. The resultant dopant concentration in the Si channel near the surface is several times lower than it is directly after the implant step above, due to segregation of Boron into gate insulator 130. As noted earlier, the Boron dopant then acts effectively as an electron trap during operation of device 100. In contrast to some of the prior art implantation techniques discussed earlier, the oxidation step appears to incorporate the Boron in a manner that facilitates shallow electron traps, making it easier for charge to move in and out of gate insulator 130.

Next, polycrystalline silicon is deposited and patterned to form gate electrode 110. N-type dopant ions such as Arsenic are subsequently implanted at moderate dose to form the lightly doped source/drain regions self-aligned to gate 110, after which sidewall spacers (not shown) are formed by conformal deposition and anisotropic etching of an insulating layer such as silicon nitride. Deep source/drain contact regions 140 and 150 are then formed by ion implantation of Arsenic or Phosphorus and thermal annealing to activate the dopants. Device fabrication is completed with standard passivation, contact and metallization processes. While not explicitly shown, it is apparent, because only conventional processing is required, that other CMOS devices can be formed in the same mask with the present NDR device 100, so that, for example, memory and logic circuits can

be formed at the same time as the present device, and thus integrated directly to form a conventional CMOS circuit having NDR capability. While the above is explained with reference to a CMOS process, it will be appreciated by those skilled in the art that other types of starting semiconductor materials could also be used instead. Suitable and/or optimal processing conditions for achieving the NDR mode in any particular CMOS compatible environment will be easily designed and determined by those skilled in the art through conventional modelling and experimentation techniques.

As a final note it is preferable that during normal operation of device 100 that a body contact (V_B) should be electrically biased (e.g. at a fixed potential of 0 Volts, as is typical for n-channel MOSFETs). If body terminal (V_B) is not connected (i.e. is "floating") then the NDR behavior is drastically diminished or even eliminated. This is because holes which are generated by hot electrons will accumulate at the channel-to-source junction, forward biasing the junction and effectively reducing the transistor threshold voltage (counteracting the charge-trapping effect of increasing the threshold voltage), if the holes are not allowed to flow out of the channel region through the body contact. Thus, if NDR-MISFET 100 is implemented in a silicon-on-insulator substrate, or in a thin film of polycrystalline silicon, care must be taken to provide a body contact. This aspect of the invention can also be exploited of course for certain applications, where it may be potentially useful to be able to turn on or turn off the NDR mode by connecting or disconnecting (switching) a bias voltage to body terminal V_B, respectively.

With the prior art, even if a device exhibiting adequate negative differential resistance can be produced, it is still a daunting task to integrate such a device into a conventional CMOS process. Since the device in this invention is inherently an NMOS structure, integration of this device with conventional logic CMOS devices is straightforward. The illustrative flow in FIG. 4 allows an NDR device process module to be completely de-coupled from a conventional process, to allow for independent optimization of the NDR devices and the CMOS devices. This makes it more straightforward to scale the NDR device in this invention with future generations of CMOS integrated-circuit technology.

Additional processing variations

Theoretical analyses performed by the inventors indicate that the following conditions and additional process details can be sufficient to achieve an NDR FET in a deep-submicron CMOS technology. In particular, our results show that a peak electric field in the channel on the order of 0.5 MV/cm or higher is preferable. Furthermore, we have also discovered that water (H_2O) or Hydrogen can be used to form the aforementioned charge traps, in addition to the other species noted earlier. In a preferred approach, water-related traps are formed by oxidation of heavily boron doped Si followed by a subsequent anneal in a steam ambient. It is believed (but has not been confirmed) that the boron assists in the formation of water-based traps.

In addition, it is further preferable, for at least some geometries, to not use structures that may inhibit the generation of hot carriers in the channel, or significantly reduce the peak electric field. For example, some forms of conventional lightly doped drain (LDD) structures can impede the generation of such carriers, particularly if they are formed prior to the channel doping noted above.

Furthermore, in the discussions above pertaining to the preferred level of p-type dopant concentration in the channel, it will be understood by those skilled in the art that this refers to a "net" p type concentration level, and not the actual absolute value that might be implanted during any particular processing step, or which might be present during any intermediate step. In other words, regardless of the intervening doping operations, the present invention can be implemented in any fashion so long as the final p-type dopant concentration in the channel is on the order of 1×10^{18} cm^{-3} or higher, because this permits an appropriate electric field strength to be generated in the channel. In fact, as is apparent from the above, the onset of the NDR behavior can be controlled as well by appropriate tailoring of the channel doping concentration as well. It should be noted that these figures are merely preferable for existing geometries, and that other suitable values will be determinable by those skilled in the art for other geometries, structures, etc., based on the present teachings and other well-known techniques.

In another variation, a preferred embodiment of the present invention can also include a counter-doping step to tailor the NDR FET threshold voltage. This step is performed after the channel doping operation noted earlier, and is simply done to reduce the net p-type concentration in the channel. A higher level of Boron (greater than 5×10^{18} cm^{-3})

may be desirable for some architectures implemented in deep submicron technologies. Conventional thermal annealing is also preferably employed to help incorporate some of the Boron into the gate dielectric so that it will facilitate the creation of appropriately configured trapping sites.

In yet another variation, although it is preferable in some substrates and applications that the body of the NDR FET be biased (e.g. at 0 V) to minimize the "floating body" effect, it is possible to tailor the design of the NDR FET to ensure that NDR behavior is maintained in the absence of a body bias. Thus, it is possible to implement the NDR FET using a silicon-on-insulator (SOI) substrate, without providing a body contact, for compact integration. Compatibility with SOI substrates is a useful feature, since such substrates will increasingly be used in IC manufacturing to achieve higher circuit operation speeds with lower power consumption, due to significant reductions in interconnect and junction capacitance.

As discussed above, a preferred primary mechanism for achieving NDR behavior in an insulated gate field-effect transistor is to trap energetic ("hot") carriers from a channel. The traps should be configured preferably so that a trap energy level should be higher than the semiconductor conduction band edge, in order for it to primarily (if not exclusively) trap hot carriers. For example, a trap which is energetically located 0.5 eV above the semiconductor conduction band edge can only trap electrons from the semiconductor which have kinetic energy equal to or greater than 0.5 eV. For high-speed NDR FET operation, it is desirable to have the carrier trapping and de-trapping processes occur as quickly as possible. As described in the above preferred embodiment, this result is achieved by placing traps in close proximity to the channel, i.e. within 1.5 nm of the gate-dielectric/semiconductor interface as previously stated.

A similar (if not superior in most cases) result would result if the traps were located right at the interface itself. In this regard it should be noted that interface traps which are energetically located well above the semiconductor conduction band edge will have no effect on FET performance until a significant percentage of the mobile carriers in the channel have sufficient kinetic energy to become trapped. The formation of such interface traps would also be preferable from a process integration standpoint, because it would eliminate the need to selectively form a separate trap-containing dielectric layer in the NDR FET regions of the semiconductor surface. Accordingly, in such instance an appropriate dopant or ion species

(of the type mentioned earlier) could be implanted/diffused to position the traps in such interface region instead. The particulars of such implantation and/or diffusion operations will vary from implementation to implementation of course based on the particular geometry, layer compositions, layer thicknesses, desired trap characteristics, desired trap locations, etc., and thus the appropriate process parameters, including ion implantation energies and species will be easily determined through routine optimization by those skilled in the art.

Another apparent observation from the present teachings is the fact that devices employing the present invention utilize a type of tunneling to a charge "trap", and not tunneling to a conduction band per se as required in conventional NDR devices such as tunnel diodes. All that is required is that the carriers be given sufficient energy to penetrate the semiconductor - insulator interface potential barrier, and then be trapped by traps within one or more dielectric layers (including any or all of the SiO_2, SiO_XN_Y and Si_3N_4 layers mentioned above). Thus, it is not necessary to set up a complicated set of precisely tuned layers in a particular fashion to achieve a continuous set of conduction bands as required in conventional NDR devices, and this is another reason why the present invention is expected to achieve more widespread use than competing technologies.

As an additional variation, the NDR FET of the present invention can be used to eliminate the need for p-channel transistors in low-power memory and logic circuits, including for example in an SRAM cell or in an inverter. In this regard, the invention provides the capability to implement low-power memory and/or logic functions using an all-NMOS (only n-channel devices) technology. Conventional CMOS technology requires significantly higher process complexity than an all-NMOS technology because of the need to define separate, electrically isolated n-type and p-type well regions, as well as separate n-type and p-type source/drain extension and contact regions. The present invention therefore provides the means to achieve more compact, simpler and overall less expensive circuit architectures and manufacturing processes.

While prior art devices (including memory cells) using single-type transistors are well-known, such devices have typically used either an active or passive load device, including for example a transistor and/or an implanted resistor or thin film resistor. The primary disadvantages of these past approaches are:

1. Significant static power dissipation
2. Reduced output voltage swing (difference between "high" and "low" values of output voltage)

Furthermore, some prior art NDR devices have been proposed as load devices for an SRAM cell (see U.S. Patent No. 6,294,412 which is incorporated by reference herein) to reduce overall cell size, but these implementations have been limited to two terminal tunneling diodes, which require a specialized sequence of processing steps and hence would increase the complexity and cost of the IC fabrication process.

In contrast, in the present invention, an NDR FET does not require any non-standard processes and only requires that an additional signal line be provided to control a gate of such FET as noted above. Furthermore, since a channel region of the NDR FET uses a dopant common to a conventional insulated gate field effect transistor (IGFET), i.e., such as an n-type channel, these features can be formed at the same time during a manufacturing process.

Accordingly, a very beneficial use of the NDR FET of the present invention would include as a substitute for the NDR devices of an SRAM memory cell of the type noted in Figures 1a and 1b of U.S. Patent No. 6,294,412 noted earlier, as either a pull-down or pull-up element in series with an appropriately biased IGFET. In such an embodiment, structural features critical to the operation of an IGFET shown therein, including for example, a gate insulation layer, source/drain regions, isolation regions, contacts, gate electrodes, etc., formed in a semiconductor die/wafer can all be formed at the same time during common operations and thus shared with an NDR FET of the present invention. From a process integration perspective, the present approach provides a substantial advantage, therefore, over mixed process load technologies.

Referring to FIG. 5, therefore, a preferred embodiment of a 2T SRAM cell 500 uses an NDR FET 510 including a gate electrode that is biased by an input signal Vbias (which can be fixed, or can be clocked). An NDR FET source is biased at a low potential Vss (e.g. 0V, or ground); its drain is connected to one of the source/drain terminals of an n-channel access transistor 520. The other source/drain terminal of access transistor 520 is connected to a data (bit) line 530. The gate of access transistor 520 is connected to a word line 540. Thus, such 2T SRAM cell requires four input/output lines: Vss, Vbias, WORD, BIT. Data is stored at node 550 shared by NDR FET 510 and access transistor 520.

To write data into cell 500, BIT line 530 is driven to an appropriate logic level (HI or LO, e.g. corresponding to a power supply Vdd or Vss, respectively), and then WORD line 540 is pulsed to a high voltage (e.g. Vdd). If the data to be written is HI, then storage node (SN) 550 is charged to the HI level through access transistor 520, and thereafter NDR FET 510 turns off. If the data to be written is LO, then SN 550 will be discharged to Vss, and NDR FET 510 turns off.

Thus, once data is written onto SN 550, the NDR FET 510 is turned off, to achieve low standby current. A pulse width (in units of time) of a WORD line voltage pulse should be wide enough to allow SN 550 to be charged fully to the HI level, or discharged fully to the LO level, and to allow NDR FET 510 to switch from NDR mode to non-NDR mode or vice versa. Accordingly the pulse width will vary of course from circuit to circuit and can be determined in accordance with well-known techniques.

To read data from cell 500, BIT line 530 is preferably precharged to the HI level, and then connected to the input of a conventional sense amplifier (not shown). WORD line 540 is then pulsed to a high voltage. If the data stored is HI, then SN 550 will not pull down the voltage on BIT line 530 through access transistor 520. If the data stored is LOW, then SN 550 will pull down the voltage on BIT line 530 through access transistor 520. As the BIT line voltage is pulled down, the voltage on SN 550 will rise, so that NDR FET 510 will turn on and help pull the BIT line voltage down through access transistor 520.

A sense amplifier can be designed through any conventional techniques to quickly detect (within nanoseconds) whether or not the BIT line voltage is being pulled down, and then amplify the data signal (e.g. outputs a voltage Vss if it detects that the BIT line voltage is dropping, otherwise maintains a high output voltage).

It should be noted that for fast data sensing (within 1 ns), a differential amplifier (requiring 2 inputs instead of 1) is preferable. In such cases a neighboring BIT line or "dummy" BIT line can be precharged to an appropriate level (e.g. Vdd/2) and used to provide the second input signal to the differential sense amplifier.

To compensate for potential leakage current and/or alpha-particle strikes, a periodic refresh can be performed to ensure that any HI voltages do not degrade with time.

It will be apparent to those skilled in the art the aforementioned NDR device can be advantageously employed in both memory and logic applications, and in the types of circuits as described in the prior art, i.e., as a memory device, as part of a logic circuit, a self-latching

logic device, an amplifier, an oscillator, power management, and many other environments where its useful characteristics can be exploited.

While this invention has been described with reference to illustrative embodiments, this description is not intended to be construed in a limiting sense. It will be clearly understood by those skilled in the art that foregoing description is merely by way of example and is not a limitation on the scope of the invention, which may be utilized in many types of integrated circuits made with conventional processing technologies. Various modifications and combinations of the illustrative embodiments, as well as other embodiments of the invention, will be apparent to persons skilled in the art upon reference to the description. Such modifications and combinations, of course, may use other features that are already known in lieu of or in addition to what is disclosed herein. It is therefore intended that the appended claims encompass any such modifications or embodiments. While such claims have been formulated based on the particular embodiments described herein, it should be apparent the scope of the disclosure herein also applies to any novel and non-obvious feature (or combination thereof) disclosed explicitly or implicitly to one of skill in the art, regardless of whether such relates to the claims as provided below, and whether or not it solves and/or mitigates all of the same technical problems described above. Finally, the applicants further reserve the right to pursue new and/or additional claims directed to any such novel and non-obvious features during the prosecution of the present application (and/or any related applications).

What is claimed is: